DOCUMENT RESUME

ED 262 051 TM 850 526

AUTHOR Biester, Thomas W.
TITLE A Study of the Relationship Between a Medical Certification Examination and an In-Training Examination.
PUB DATE Apr 85
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985). Appendices contain broken type.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Certification; Correlation; Department Heads; Educational Assessment; Expectancy Tables; *Graduate Medical Education; Graduate Medical Students; Higher Education; *Occupational Tests; Predictive Validity; Psychometrics; Regression (Statistics); *Surgery; Testing Problems; Testing Programs; *Test Interpretation; *Test Use
IDENTIFIERS *American Board of Surgery In Training Examination; American Board of Surgery Qualifying Examination

ABSTRACT

This paper presents results of a study examining the relationship between the American Board of Surgery's In-Training Examination and its Qualifying Examination. 1982 Test scores on both examinations of 764 candidates in their fifth year of training as Chief Residents were analyzed. Descriptive statistics, correlations, and standard errors were computed. A moderately high correlation coefficient was observed. The findings do not support those previously reported in the literature. Regression equations and theoretical expectancy tables were prepared for use by residency program directors in predicting future success in the certification process. Various factors that possibly influenced the size of the correlation coefficient are discussed, including examination purpose/content, security breaches, motivations to take the examination, and other factors. Implications for other medical specialties in analyzing and interpreting In-Training Examination results are discussed. (Author/BS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED262051

A STUDY OF THE RELATIONSHIP BETWEEN A MEDICAL CERTIFICATION EXAMINATION AND AN IN-TRAINING EXAMINATION

THOMAS W. BIESTER

AMERICAN BOARD OF SURGERY
1617 JOHN F. KENNEDY BLVD.
SUITE 1561
PHILADELPHIA, PA 19103

A PAPER PRESENTED AT THE ANNUAL MEETING OF THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, CHICAGO, ILLINOIS, APRIL, 1985

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Biester, Thomas W.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 850 526

ABSTRACT

A Study of the Relationship Between a Medical Certification Examination and an In-Training Examination.

Thomas W. Biester, American Board of Surgery

This paper presents results of a study examining the relationship between the American Board of Surgery's In-Training Examination and Qualifying Examination. A moderately high correlation coefficient was observed. The findings do not support those previously reported in the literature. Regression equations and theoretical expectancy tables were prepared for use by residency program directors in predicting future success in the certification process. Various factors that possibly influenced the size of the correlation coefficient are discussed, including examination purpose/content, security breaches, motivations to take the examination, and other factors. Implications for other medical specialties in analyzing and interpreting In-Training Examination results are discussed.

INTRODUCTION

Similar to guidelines for other medical specialties, the "Essentials for Approved Residencies" in Surgery require that programs provide a clear demonstration of each resident's competence in basic surgical knowledge before progressing to the next level of supervised semi-independent operative experience and patient management. When these guidelines were first published in 1974, program directors in surgery urged the American Board of Surgery (ABS) to develop a written examination for the assessment of surgical residents. The request came at a time when the Board was becoming concerned about the relatively high failure rates on its certification examinations and about what it perceived as deficits in the cognitive knowledge base of many residents. The first In-Training Examination (ITE) developed by the Board was administered in 1975. This paper examines the use of the ITE for resident evaluation after ten years in the field and, in particular, the relationship between ITE results and performance on the Board's Qualifying Examination (QE), the written examination (Part I) necessary for certification in Surgery.

BACKGROUND ON IN-TRAINING EXAMINATIONS

In-Training Examinations sponsored by medical specialty organizations date back to the early sixties in orthopaedic surgery and neurological surgery (Levit, 1969). The specialty boards for these two disciplines were concerned about the high failure rate of candidates appearing for certification, despite the uniformity of the prescribed period of training. The 1964 In-Training Examination for these specialties was developed with the assistance of the National Board of Medical Examiners.

In-Training Examinations are generally oriented toward three primary goals: 1) to define a content domain of knowledge necessary for competent practice; 2) to assess the relative strengths and weaknesses of individual residents at a time early enough in their training that deficits can be corrected; and, 3) to help directors of residency programs detect areas of relative strengths and weaknesses of their program (Grosse, Cruft, and Blaisdell, 1980). It is important to keep these goals in mind. The first ABS ITE was welcomed by program directors as a valuable formative tool (Friedmann, 1985).

The American Board of Medical Specialties reports that written In-Training Examinations are now offered in 17 of the 23 recognized medical specialties, with three others currently in the planning stage (ABMS, 1984). Table 1 lists the specialties that now administer ITEs. Seven of these examinations are sponsored by the specialty board, with the others sponsored by a professional society, academy, or college. All are administered annually, and all are open to residents, with many also open to practitioners. Many, such as the ABS ITE, are designed primarily for use by the residency program director, rather than the residents themselves. Internal Medicine is a noteworthy exception to the specialties offering written ITEs, instead requiring a detailed resident evaluation system as a key element of the credentialing process. In some cases (e.g., Surgery), ITEs are developed through the traditional committee approach while in others (e.g., Pediatrics), examinations are comprised of items used in other examinations, such as prior certifying examinations. Specialty In-Training Examinations are distinguished from the self-assessment programs/examinations offered by the professional societies in most specialties. While the overall goal is diagnosis and self-improvement, the

TABLE 1

MEDICAL SPECIALTIES OFFERING IN-TRAINING EXAMINATIONS
AND/OR SELF-ASSESSMENT PROGRAMS/EXAMINATIONS

Specialty	In-Training Examination	Sponsor*	Self-Assessment Program/Examination
Allergy & Immunology	-	-	-
Anesthesiology	yes	B & S	yes
Colon & Rectal Surgery	-	-	yes
Dermatology	yes	[illegible]	yes
Emergency Medicine	-	-	yes
Family Practice	yes	B	yes
Internal Medicine	-	-	yes
Neurological Surgery	yes	B	yes
Nuclear Medicine	-	-	yes
Obstetrics & Gynecology	yes	S	yes
Opthalmology	yes	S	yes
Orthopaedic Surgery	yes	S	yes**
Otolaryngology	yes	S	yes
Pathology	-	-	yes
Pediatrics	yes	B	yes
Physical Medicine & Rehabilitation	yes	S	yes
Plastic Surgery	yes	S	yes
Preventive Medicine	-	-	-
Psychiatry & Neurology	yes	S	yes
Radiology	yes	S	yes
Surgery	yes	B	yes
Thoracic Surgery	yes	B	-
Urology	yes	S	yes

* Sponsor Key: B = Board
S = Professional Society, College, Academy

** Not open to residents

Source: American Board of Medical Specialties 1984 Annual Report

latter generally lack standard administration conditions and national score norms. They are offered to individuals rather than residency program directors and are usually targeted at a knowledge level somewhat beyond that of a junior resident.

IN-TRAINING EXAMINATIONS AND THE CERTIFICATION PROCESS

The three primary goals listed above illustrate the basic formative nature of In-Training Examinations, namely diagnosis of weaknesses for self-improvement prior to completion of graduate medical training. Except for two specialties (Anesthesiology and Neurological Surgery), the ITEs are not an official part of the credentialing process. However, in no case is an ITE officially used to evaluate a program and results are not reported to Residency Review Committees (RRCs), unless the program director voluntarily elects to do so. Still, since RRCs for most medical specialties use the number of residents who successfully complete certification requirements for the particular specialty board as one important criterion for program accreditation, residency directors, responsible for assuring the professional competence of their graduates, have often expressed an interest in knowing the relationship between In-Training Examinations and board Certifying Examinations. If the relationship is strong and both examinations are content valid, the ITE is a useful tool for program directors in determining that their residents have acquired the basic knowledge needed for competent practice in the field. Careful monitoring of progress, through ITEs and other relevant assessment procedures throughout the course of residency, helps to ensure that graduates are knowledgeable in their field and certification success rates, as

validation criteria, will be high. More important, graduates will be competent practitioners.

Shetler (1982) investigated the relationship between the ABS ITE and the ABS QE, the initial step in the certification process. This study looked at results from one particular residency program. He reported a correlation of .79 and suggested guidelines for program directors on how to use and interpret ITE results. However, the usefulness of this study's findings is somewhat limited because of the small sample size and some other technical flaws (e.g., use of percentile scores as a basis for calculations).

Garvin and Kaminsky (1984) followed Shetler's model with residents from their program over a five-year period. Their findings were virtually identical to Shetler's, with a correlation of .76. Results of both of these studies received widespread circulation and are frequently cited by program directors in surgery. However, since some questioned the validity of findings and because of the small sample size, the ABS decided to conduct further investigations of the relationship between its In-Training and Qualifying Examinations.

STUDY OBJECTIVES

The current study was designed to enhance the findings of the Shetler and Garvin-Kaminsky studies by examining ITE and QE results for all surgery residents participating in the ABS assessment program. Specifically, three major objectives or research questions were addressed:

1) What is the observed relationship between the American Board of Surgery's In-Training Examination and Qualifying Examination?

2) What factors seem to influence the strength of the relationship between the ITE and the QE?

3) What are the implications of the relationship for residency program directors in surgery?

Although the objectives are essentially the same as those addressed by the reported studies, findings are more generalized and accurate. The findings and methodologies are also somewhat generalizable to other medical specialty areas that administer ITEs, suggesting a fourth objective:

4) What are the implications of the study for other medical specialty areas?

METHODS AND DATA SOURCES

In 1982, the ABS ITE was administered to 7349 residents, representing virtually all of the accredited surgery residency programs nationally. Of this number, 915 were in their fifth year of training as Chief Residents. A total of 764 of these Level V candidates also took the ABS QE in 1982. Test scores for this group on both examinations were analyzed. Standard scores, based on a mean of 500 and a standard deviation of 100, were used as a basis for the analysis. Descriptive statistics, correlations, and standard errors were computed. A scatter plot was prepared and a regression equation for predicting an individual's QE score on the basis of the ITE score was constructed. Finally, a theoretical expectancy chart showing the probability of passing the QE at different level score intervals on the ITE was prepared.

RESULTS

Prior to formal data analysis, the psychometric properties of both examinations were considered. The 1982 ITE had 351 scorable units (although 200 items, many had several scorable units), with an average difficulty or "p"

value of .72 and overall reliability of .97. The 1982 QE had 770 scorable units (302 items), with an average "p" of .74 and overall reliability of .99. The high reliabilities indicate that the precision of both examinations is exceptionally good and one should be fairly confident about individuals' scores.

The means for the study group were 589 and 557 for the ITE and QE, respectively. The high standard score for the ITE was as expected, since Level V scores are much higher than the total group taking the ITE. The high mean score for the QE indicates that the study group is not representative of the total group taking the examination, but is fairly representative of the group of U.S./Canadian medical school graduates taking the test for the first time. The score ranges were 370-732 for the ITE and 225-715 for the QE. The ITE range was somewhat restricted, as expected, because of the nature of the resident group.

The observed correlation coefficient between the two examinations was .48. This is contrasted to coefficients of .79 and .76 found in earlier studies. Thus, while these authors concluded that the correlation between the ABS ITE and QE was very high, the conclusion drawn from the Board study is that it is only moderately high (only 23% of the variance is explained). The regression equation for predicting the QE score on the basis of the ITE score is as follows:

$$\widehat{QE} = .575 \text{ ITE (V)} + 218.15.$$

However, the standard error of estimate is fairly high at about 50 points. Thus, the predicted score range would be at least 100 points, probably not that useful for program directors.

Most studies of this type reported in the literature use the regression approach for prediction. However, as Guion (1965) noted in his Personnel Testing textbook, such an approach is not most appropriate for making categorical predictions. Such is the case when trying to predict passing or failing the Qualifying or other certification examination on the basis of In-Training Examination results. For this purpose, Guion recommends theoretical expectancy charts. Table 2 shows the QE pass/fail expectancies for the ABS given ITE scores at various intervals. This chart would lead one to choose a score of about 490-510 for a Level V resident as the basis for predicting the QE. Even with a moderate correlation, one could correctly predict passing or failing the QE at least 85% of the time using this chart. Such a chart would be very useful for directors of surgery residency programs. In addition, such an approach, although fairly simple, would be useful for all residency programs. Directors should also consider typical score trends in assessing growth through each level of the residency program. If scores at lower levels are below that for which a passing score at the Senior Resident level would be predicted, appropriate remedial education could be prescribed.

DISCUSSION

The correlation results are disappointingly low, particularly in light of the high coefficients observed in the earlier studies. One major reason for the fairly low observed correlation is the probability of security breaches in the administration of the ITE. For example, it has been reported that some program directors may allow open-book administration or proctors may not strictly adhere to test administration guidelines. In addition, for some time it has been clear that copies of prior ITEs have been reproduced and

TABLE 2

EXPECTANCY CHART FOR PREDICTING PASS/FAIL ON ABS QUALIFYING EXAM

ON BASIS OF ABS ITE (LEVEL V) SCORE*

ITE Score Range (Level V)	# of Examinees (1982)**	Proportion Passing QE (@495)***	Proportion Failing QE (@495)
711 +	5	1.00	0.00
671 - 710	42	.93	.07
631 - 670	142	.92	.08
591 - 630	218	.90	.10
551 - 590	181	.86	.14
511 - 550	110	.85	.15
471 - 510	39	.46	.54
431 - 470	19	.37	.63
- 430	8	.25	.75
TOTAL	764	.84	.16

Observed correlation between ITE (Level V) and QE = .48.

Regression equation for predicting QE score from Level C ITE score:

$\widehat{QE} = .575 \text{ ITE (V)} + 218.15$.

Standard error of estimate = 50.45; thus, predicted QE should be in the range of QE ± 50.45 most of the time.

* Based on empirical results from 1982.

** Study group includes all examinees who took both ITE and QE in 1982.

*** Passing score of 495 in 1982 represents 1.08 standard deviation below the mean of the reference group.

distributed throughout many parts of the country. Since previously used questions comprised about 57% of the ITE content in 1982, results might not accurately represent the true knowledge of all Level V candidates. Indeed, a recent ABS study of performance on new versus used items on the 1984 ITE indicated "suspicious" scores for about 20% of the surgical residency programs. This would introduce a large error factor into the score results and the correlation would be significantly reduced. Analyses showed that ITE results for residents in these programs could be inflated by 50 to 125 standard score points. Security measures were taken for the 1985 examination and a follow-up study examining results for selected programs with no apparent security problems is currently underway in order to gain a more valid picture of the relationship between the ITE and the QE.

Residency programs for other specialties should likewise be aware of the problems that breaches of examination security pose for accurately interpreting ITE results. Friedmann (1985) indicated that seven specialties, including surgery, reported security problems associated with their ITEs. Although many program directors become incensed at charges of cheating on the part of residents, Ballinger (1985) and Friedmann provide good historical overviews that help explain why such incidents may be increasing. Although their focus is on the surgery examinations, their comments apply as well to other specialties. In earlier years (e.g., early to mid seventies), not much importance was attached to the ITEs; they were used primarily in a formative sense to diagnose weaknesses in the residents' knowledge base. Times changed, however, and the examinations, in some instances, have become pivotal in the lives of residents, particularly in pyramidal programs, who are competing for fewer positions, in fewer programs, where career success may hinge on test

outcomes. In addition, as the number of programs dwindled, program directors became concerned that they would be judged (e.g., by the RRC) on how their residents performed on the ITE. Clearly, the attitudes and motivations of program directors and residents strongly influence performance on the examination.

Berry (1984) noted various uses and abuses of the ABS ITE, including the increasing shift from formative to summative interpretations. This shift has led to a great deal of confusion and anxiety. A Board survey of program directors indicated that 34% used a formal cutting score for determining passing or failing the ITE. Most (85%) noted that they consider test results in resident promotion decisions. Five other Boards reported setting formal passing scores for their ITEs. Although 84% of the surgery program directors indicated that they use the ITE for formative evaluation of their residents, the three primary goals apparently need to be reemphasized.

While ABS ITE results are reported in a norm-referenced sense (e.g., standard scores and national percentiles), results by program and individual resident are also reported in a criterion-referenced sense. Keywords that describe the specific items answered incorrectly by each examinee and the entire group of examinees are reported to program directors. An example is included in the Appendix. These results can be used for overall program improvement and individual remediation. In this regard, Dean, Hanni, Pyle, and Nicholas (1984) reported positive effects of coaching and remedial programmed instruction on ABS ITE scores. Program directors and residents alike have repeatedly requested that copies of previous examinations be made available for self study. However, this would create the need for new test forms each

year, would probably cause some equating and norming problems, and would be very expensive. In addition, the SESAP examinaitons of the American College of Surgeons serve this purpose.

Another possible cause for the low correlation, somewhat related to the above discussion, is the varying amount of preparation that residents make for the examination. Since residents differ with regard to their motivation to succeed on the examination, some prepare extensively for it while others do not. An ABS survey of chief residents indicated that 51% actively studied for the ITE prior to administration. While Donovan (1985) notes that "cramming" should not significantly affect performance, the results of the study by Dean and associates may suggest otherwise. Some program directors have indicated that formal study groups have successful payoffs while others prefer not to encourage special cramming but rather to assess the level of knowledge retained throughout the normal course of the residency (in other words, use the ITE as a spot check).

Both the ABS ITE and QE are designed to assess examinees' level of knowledge related to surgery, in order to assure that all residents completing training possess the essential or "core" knowledge necessary for competent practice. As such, one might expect correlations to be high. Although each examination is constructed by a different committee, using a slightly different content outline, the basic content categories are fairly similar. However, as Donovan (1985) notes, the distribution of questions on the ABS ITE is very different from the QE. The former is aimed at residents completing their second year of graduate training while the latter is appropriate for someone who has just finished the fifth year of residency. A large part of the ITE is oriented

toward factual recall of basic science concepts rather than the evaluation of judgment and patient management (not PMPs, however) emphasized in the QE. A frequent criticism of the ITE from program directors and residents is that it should contain more patient management items. There are some other content differences, as well. For example, the ITE has a heavy emphasis on knowledge of "the body as a whole" while the QE deals more with diagnosis and treatment of more specific organ systems. The audience for each particular examination is important to keep in mind, since surgery residents gradually acquire more intensive clinical experience in their fourth and fifth years. Clarke (1982) documented increasing decision-making skills from junior to chief-residency levels in surgery.

Studies by the American Board of Internal Medicine found that multiple-choice questions assessing knowledge, synthesis, and clinical judgment measure essentially the same aspects of clinical competence (Norcini, Swanson, Webster, and Grosse, 1984). However, these studies were conducted within the same examinee groups, unlike the case described above for surgery where the target audiences and corresponding hierarchial level of cognitive achievement differ.

Other studies that have examined the relationship between knowledge as measured by multiple-choice examinations and other measures of clinical competence may shed some light on the observed ITE-QE correlations. Lazar, DeLand, and Tompkins (1980) studied the relationship between ABS ITE scores and other measures of clinical competence, finding a very low correlation (.27). They concluded that this low correlation was desirable since the indicators were measuring different aspects of resident competence. Similarly

low correlations between multiple-choice test scores and clinical competence ratings were reported by Stillman (1984) and Berenson, Stimmel, and Aufses (1981). In the Board survey of program directors, 58% reported that the ITE results correlated well with their impressions of their residents' clinical performance.

From these studies and the observed ITE-QE correlation one might conclude that the examinations are, in fact, measuring different, but somewhat related attributes. Clarke (1982) stated that the satisfactory surgery resident requires both factual knowledge and judgment as well as the ability to combine these assets into clinical decision making. The ITE, while providing a valid assessment of residents' requisite factual knowledge (as indicated by 96% of surveyed program directors), probably does not adequately tap judgment and clinical decision-making. As such, Lazar and associates emphasized that the ITE cannot be used as a sole basis for assessing residents' clinical competence. Shetler and Berry echo this sentiment, and, likewise, the Board has taken a strong position encouraging program directors to use multiple criteria in assessing the competence of their residents.

Finally, various statistical factors may contribute to the relatively low correlation. These include restriction of range and probable ceiling effects.

SUMMARY AND CONCLUSIONS

Previously reported findings concerning the degree of the relationship between the ABS ITE and QE were not observed. The actual correlation was only moderately high. Apparently, the two examinations are measuring somewhat

different attributes. Other factors that possibly account for the lower correlation are security problems, motivation to take the examinations, uses and abuses of results by program directors, the nature of residency training, and statistical factors. Since the philosophy and general content guidelines of ABS' ITEs are similar to the In-Training examinations of other specialty boards, results are generalizable to the other medical specialties, as well. Factors affecting the scores on the ABS ITE probably also affect ITE scores for other medical specialties. In addition, the study provides these specialties with a useful methodology for exploring the relationship between their in-training assessments and certification examinations. In particular, the following implications can be drawn from the study:

- Program directors need to keep the basic goals of In-Training Examinations in mind. In the formative sense, results can be used to direct remedial study for residents and programmatic deficiencies can be pinpointed and corrected. Further development of criterion-referenced reporting approaches is encouraged.

- While the assessment of residents' cognitive knowledge is important, program directors need to keep in mind that other measures of competence are likewise important. These other attributes are probably not assessed in typical written In-Training Examinations. New tools for the assessment of resident competence need to be developed where current state-of-the-art is inadequate.

- Program directors should be sensitive to the possible abuses of In-Training Examination results. They should also recognize that results are primarily for their use and Residency Review Committees do not require actual scores as part of the accreditation process.

- Program directors should be sensitive to security issues involved in the administration of the In-Training Examinations. To the extent that security problems exist, scores and national percentiles may be somewhat inaccurate.

- The regression approach for predicting outcomes such as specialty certification may not be the most useful or appropriate one for program directors. Rather, an approach using theoretical expectancy charts is recommended.

- Further studies of trends in the growth of residents' knowledge throughout the course of their training need to be conducted.

In-Training Examination results provide a useful tool for monitoring the quality of graduate medical education and assuring that residents acquire the essential knowledge needed to practice their chosen medical specialty. As a tool, the examination must be used properly. Test scores provide an important piece of information about the resident, but that information needs to be considered along with many other data sources concerning the residents' knowledge, skills, attitudes, and other behavior. As Shetler noted, In-Training Examinations will never completely replace the discerning intuition of a sagacious teaching staff committed to their residents, their program, and the medical discipline.

REFERENCES

American Board of Medical Specialties. Annual Report and Reference Handbook 1984. Evanston, IL: Author, 1984.

Ballinger, W. F. "The validity and uses of the In-Training Examination." Bulletin of the American College of Surgeons, 70 (1985), 12-76.

Berenson, T. F., Stimmel, B., and Aufses, A. "Concordance of surgical clerkship performances and the National Board of Medical Examiners Part II Subtest scores. A validation model." Surgery, 89 (1981), 692-696.

Berry, R. E. "ABS ITE: uses and abuses." Current Surgery, 41 (1984), 329-334.

Clarke, J. R. "The role of decision skills and medical knowledge in the clinical judgment of surgical residents." Surgery, 92 (1982), 153-158.

Dean, R. E., Hanni, C. L., Pyle, M. J., and Nicholas, W. R. "Influence of programmed textbook review on American Board of Surgery In-Service Examination scores." American Surgeon, 50 (1984), 345-349.

Donovan, A. J. "An analysis of results of the In-Training Examination." Bulletin of the American College of Surgeons, 70 (1985), 3-6.

Friedmann, P. "A program director's view of the In-Training Examination." Bulletin of the American College of Surgeons, 70 (1985), 7-11.

Garvin, P. J. and Kaminsky, D. L. "Significance of the in-training examination in a surgical residency program." Surgery, 96 (1984), 109-113.

Grosse, M. E., Cruft, G. E., and Blaisdell, F. W. "The American Board of Surgery In-Training Examination." Archives of Surgery, 115 (1980), 654-657.

Guion, R. M. Personnel testing. New York: McGraw-Hill, 1965.

Lazar, H. L., DeLand, E. C., and Tompkins, R. K. "Clinical performance versus in-training examinations as measures of surgical competence." Surgery, 87 (1980), 357-362.

Levit, E. J. "In-training evaluation of learning: objective measurement of the process and product of graduate medical education." Archives of Dermatology, 99 (1969), 242-249.

Norcini, J. J., Swanson, D. B., Grosso, L. J., and Webster, G. D. "A comparison of knowledge, synthesis, and clinical judgment multiple-choice questions in the assessment of physician competence." Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans, April, 1984.

Shetler, P. L. "Observations on the American Board of Surgery In-Training

Examination, Board results, and conference attendance." American Journal of Surgery, 144 (1982), 292-294.

Stillman, R. M. "Pitfalls in evaluating the surgical student." Surgery, 96 (1984), 92-96.

APPENDIX

AMERICAN BOARD OF SURGERY 1984 IN-TRAINING EXAMINATION
REPORT OF PERFORMANCE

PROGRAM DESIGNATION:
CODE NUMBER: 000

THE EXAMINATION WAS EQUALLY DIVIDED INTO FIVE MAJOR CATEGORIES:

A. BODY AS A WHOLE
B. GASTROINTESTINAL SYSTEM
C. CARDIOVASCULAR AND RESPIRATORY SYSTEMS
D. GU, HEAD AND NECK, MUSCULOSKELETAL, ETC.
E. ENDOCRINE, HEMATIC, LYMPHATIC SYSTEMS, BREAST

SCORES AND PERCENTILE RANK OF LEVEL I EXAMINEES:

NAME	TOTAL-%ILE	A-%ILE	B-%ILE	C-%ILE	D-%ILE	E-%ILE
	324-10	343-13	432-54	323-10	408-31	251-01
	341-16	302-07	316-12	354-20	383-25	477-67
	480-75	377-25	454-63	546-87	490-63	529-82
	354-23	435-49	355-23	384-26	357-19	366-19
	357-23	349-13	450-54	349-14	327-09	414-42

SCORES AND PERCENTILE RANK OF LEVEL II EXAMINEES:

NAME	TOTAL-%ILE	A-%ILE	B-%ILE	C-%ILE	D-%ILE	E-%ILE
	340-05	331-06	295-03	349-04	449-28	414-25
	448-31	377-11	368-09	428-25	617-91	500-54
	403-17	366-08	463-40	380-11	490-44	390-12
	441-31	401-16	467-40	467-41	525-60	386-12

SCORES AND PERCENTILE RANK OF LEVEL III EXAMINEES:

NAME	TOTAL-%ILE	A-%ILE	B-%ILE	C-%ILE	D-%ILE	E-%ILE
	441-10	441-13	411-06	498-28	525-46	381-04
	547-50	545-52	553-50	520-34	449-15	625-88

SCORES AND PERCENTILE RANK OF LEVEL IV EXAMINEES:

NAME	TOTAL-%ILE	A-%ILE	B-%ILE	C-%ILE	D-%ILE	E-%ILE
	678-92	614-71	695-97	703-96	617-74	596-59
	448-04	441-09	536-26	428-05	423-07	443-07

SCORES AND PERCENTILE RANK OF LEVEL V EXAMINEES:

NAME	TOTAL-%ILE	A-%ILE	B-%ILE	C-%ILE	D-%ILE	E-%ILE
	645-77	649-80	596-56	642-78	561-43	659-86
	675-93	631-80	644-80	682-91	658-88	611-67

EACH SCORE IS REPORTED IN STANDARD SCORE UNITS. THE AVERAGE OF THE FIVE SUB-SCORES DOES NOT EQUAL THE TOTAL TEST SCORE IN EVERY CASE BECAUSE OF ROUNDING OF SUB-TEST (CATEGORY) SCORES.

PLEASE REFER TO THE ENCLOSED BOOKLET OF NORM TABLES TO INTERPRET THESE SCORES IN RELATION TO THE PERFORMANCE OF THE VARIOUS LEVELS OF EXAMINEES.

TOTAL EXAMINEES = 15

BEST COPY AVAILABLE

1984 AMERICAN BOARD OF SURGERY IN-TRAINING EXAMINATION

PROGRAM SUMMARY OF RESIDENTS' PERFORMANCE

LISTED BELOW ARE THE ITEM NUMBERS AND KEYWORD PHRASES WHICH IDENTIFY THE QUESTION IN THE EXAMINATION WHICH WERE ANSWERED INCORRECTLY BY ONE OR MORE OF THE RESIDENT IN YOUR PROGRAM. THE NUMBER OF RESIDENTS WHO ANSWERED ANY QUESTION INCORRECTLY APPEARS IN THE APPROPRIATE COLUMNS TO THE RIGHT OF THE KEYWORD PHRASE. IF ALL RESIDENTS ANSWERED CORRECTLY, A ZERO APPEARS.

PROGRAM DESIGNATION: [illegible] IERS

CODE NUMBER: 00C

		NUMBER OF RESIDENTS TAKING THE EXAMINATION					
		LVL-1 N=005	LVL-2 N=004	LVL-3 N=002	LVL-4 N=002	LVL-5 N=002	LVL- N=00
ITEM		NUMBER OF RESIDENTS ANSWERING INCORRECTLY					
	BODY AS A WHOLE						
1	DX TPN ASSOC HYPOGLYCEMIA	1	3	1	0	0	0
[illegible]c	CHARACT INCARCER INGUINAL HERNIA	2	2	2	0	1	0
[illegible]7	RX REACTION LOCAL ANESTHESIA	3	1	2	1	1	0
[illegible]3	COMPLIC JEJUNOSTOMY TUBE FEEDING	1	0	1	0	0	0
[illegible]29	RX HYPOCHLOREMIC ALKALOSIS	4	3	1	2	0	0
[illegible]31	PREVENTION SURG WOUND INFECT	2	1	1	1	0	0
36	DX MEASURE CA BREAST	0	1	0	0	0	0
[illegible]5	RX WOUND DEHISCENCE	4	4	1	2	0	0
[illegible]3	SIGNS TRANSFUSION REACTION	3	2	0	1	0	0
55	FLUID REQUIR BURN WOUND	3	1	0	0	0	0
[illegible]6	RX BURN WOUND	3	0	0	0	0	0
[illegible]73	IMPT BLOOD ELEM WOUND HEALING	3	3	1	1	1	0
[illegible]0	PROGNOSIS SARCOMA LOWER EXTREM	2	2	0	0	0	0
81	COMPLIC SWAN-GANZ CATHETER	3	1	0	0	1	0
[illegible]82	HISTOPATHOLOGY REJECTED TISSUE	4	3	0	0	1	0
[illegible]88	ETIOL ECTHYMA GANGRENOSUM	2	4	1	1	0	0
[illegible]01	PHYSIOL ANTILYMPHOCYTE GLOBULIN	4	2	0	1	0	0
[illegible]6	RATIONALE LOW DOSE ASPIRIN RX	0	0	0	0	0	0
[illegible]1	DX STUDY CLASSICAL HEMOPHILIA	3	4	2	1	0	0
[illegible]7	MOST EFFICIENT ENERGY SOURCE	0	0	0	1	0	0
[illegible]5	ETIOL CHR ABDOMINAL SINUS	1	0	0	0	0	0
[illegible]9	ETIOL HYPERGLYCEMIA/TPN	3	3	0	0	0	0
[illegible]5	RX RESISTANT CARDIOGEN SHOCK	0	1	0	0	0	0
[illegible]4	COND ASSOC TT PURPURA	3	4	2	1	0	0
[illegible]3	URGENT RX INFANT HERNIAS	0	0	1	0	0	0
[illegible]4	RX GUNSHOT WOUND CHEST	5	4	1	2	2	0
7	CHARACT 2ND PHASE WOUND HEALING	5	3	2	2	2	0
[illegible]9	COND ASSOC HYPERCALCEMIA	0	0	0	0	0	0
[illegible]2	BACTER ETIOL SYNERGIST GANGRENE	4	4	1	0	0	0

BEST COPY AVAILABLE

1984 AMERICAN BOARD OF SURGERY IN-TRAINING EXAMINATION

ITEM PERFORMANCE REPORT

NAME: ID NUMBER: 000

SCORES AND PERCENTILE RANKING (RANKING WITHIN LEVEL OF TRAINING)

CATEGORY	SCORE	%ILE
BODY AS A WHOLE	545	52
GASTROINTESTINAL SYSTEM	553	50
CARDIOVASCULAR AND RESPIRATORY SYSTEMS	520	34
GU, HEAD AND NECK, MUSCULOSKELETAL, ETC.	449	15
ENDOCRINE, HEMATIC, LYMPHATIC SYSTEMS, BREAST	625	88
TOTAL TEST	547	50

LISTED BELOW ARE THE KEYWORD PHRASES WHICH IDENTIFY THE QUESTIONS IN EACH CATEGORY WHICH THE EXAMINEE ANSWERED INCORRECTLY. A KEYWORD PHRASE IS LISTED FOR EACH BEST-ANSWER QUESTION ANSWERED INCORRECTLY AND FOR EACH MULTIPLE TRUE-FALSE QUESTION IN WHICH AT LEAST TWO PARTS WERE ANSWERED INCORRECTLY.

BODY AS A WHOLE

- TPN ASSOC HYPOGLYCEMIA
- REACTION LOCAL ANESTHESIA
- IDL ECTHYMA GANGRENOSUM
- AND ASSOC TT PURPURA
- EP ANTICOAG PT FOR APPENDECT
- CHARACT INCARCER INGUINAL HERNIA
- IMPT BLOOD ELEM WOUND HEALING
- DX STUDY CLASSICAL HEMOPHILIA
- CHARACT 2ND PHASE WOUND HEALING
- RX AMEBIC ABSCESS LIVER

GASTROINTESTINAL SYSTEM

- STUDY BLUNT ABD TRAUMA
- TC CARCINOIDS
- ST COMMON CONG ANOM ESOPH
- ARACT VASCULAR ECTASIAS LG BOWEL
- ARACT DISTAL SPLENORENAL SHUNT
- RX DRUG COLON BACT INFECTION
- PHYSIOL ACINAR CELLS PANCREAS
- MALIGNANCY/CHR ULCER COLITIS
- USE MESOCAVAL H-GRAFT

CARDIOVASCULAR AND RESPIRATORY SYSTEMS

- AT RT HEPATIC ARTERY
- RESP DISTRES POSTOP RLLOBECT
- FINITION TOTAL LUNG CAPACITY
- ARACT BRONCHIAL ADENOMAS
- CLAUDICATION
- IOL AORTIC VALVE DIS
- ARACT FUNCT RESID CAPAC
- DX SOURCE PLEURAL FLUID
- COND ASSOC ASD (SV TYPE)
- CHARACT CARDIAC MYXOMA
- DX ALVEOLAR HYPOVENTILATION
- DX BOECK'S SARCOID
- LONGEVITY POST CARDIOVASC PROCED
- CHARACT CARDIAC VALVE SUBSTITUTES

GU, HEAD AND NECK, MUSCULOSKELETAL, ETC.

- INCREASE INTRACRAN PRESS
- ACUTE PARONYCHIA
- OSIS DIGITAL NERVE REPAIR
- ST COMMON SITE KNEE INJURY
- CA BLADDER
- NERVE INJ SHOULDER DISLOC
- RX SKULL FX, OTORRHEA
- RX BOWEN'S DIS CERVIX
- TYPE NEUROGENIC BLADDER
- RX BARTHOLIN'S ABSCESS

BEST COPY AVAILABLE